

UNCLASSIFIED

Distribution A: Approved for public release; distribution is unlimited.

(U) Active Wiki Knowledge Repository

October 2012

Scott C. McGirr, Ronald J. Wroblewski, and Eric L. Dorman

Space and Naval Warfare Systems Center Pacific
53360 Hull Street, San Diego, CA 92152-5001, USA

scott.mcgirr@navy.mil, ronald.wroblewski1@navy.mil, eric.dorman@navy.mil

ABSTRACT

An Active-Wiki concept was designed and a prototype built to evaluate a dynamic customized knowledge repository. By its nature, a wiki framework enables distributed sharing of information. The capability was designed to enable users to build and share custom views of information. The wiki was termed "active" because it automatically monitors information sources, updates existing pages with content or hyperlinks and, when needed, builds new pages. The Active Wiki (AW) maintains page relationships based on a user's knowledge-domain ontology. The prototype facilitates knowledge formation by providing users tools to analyze and visualize content. The novice user can use existing pages and/or tailor pages through menu selections. The advanced user can customize content and tailor views by writing Ruby scripts. A powerful aspect of the Active Wiki is that visualizations of information are not end points but rather process outputs that can be manipulated and used as input to other processes. To handle big data efficiently, the design concept uses metadata geo-hashing methods for data referencing and searching files. Open layers are used for overlaying content for virtual data fusion. The prototype uses modular software development to lower costs and open standards for construction pages to improve interoperability. The AW has been demonstrated and tested in a number of exercises.

Keywords: wiki; knowledge base; RDF; triple store; data fusion; cloud computing

UNCLASSIFIED

Report Documentation Page				Form Approved OMB No. 0704-0188		
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.						
1. REPORT DATE OCT 2012		2. REPORT TYPE N/A		3. DATES COVERED -		
4. TITLE AND SUBTITLE Active Wiki Knowledge Repository				5a. CONTRACT NUMBER		
				5b. GRANT NUMBER		
				5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S)				5d. PROJECT NUMBER		
				5e. TASK NUMBER		
				5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Space and Naval Warfare Systems Center Pacific 53360 Hull Street, San Diego, CA 92152-5001, USA				8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)		
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited						
13. SUPPLEMENTARY NOTES See also ADM202976. 2012 Joint Meeting of the Military Sensing Symposia (MSS) held in Washington, DC on October 22-25, 2012.						
14. ABSTRACT An Active-Wiki concept was designed and a prototype built to evaluate a dynamic customized knowledge repository. By its nature, a wiki framework enables distributed sharing of information. The capability was designed to enable users to build and share custom views of information. The wiki was termed "active" because it automatically monitors information sources, updates existing pages with content or hyperlinks and, when needed, builds new pages. The Active Wiki (AW) maintains page relationships based on a users knowledge-domain ontology. The prototype facilitates knowledge formation by providing users tools to analyze and visualize content. The novice user can use existing pages and/or tailor pages through menu selections. The advanced user can customize content and tailor views by writing Ruby scripts. A powerful aspect of the Active Wiki is that visualizations of information are not end points but rather process outputs that can be manipulated and used as input to other processes. To handle big data efficiently, the design concept uses metadata geo-hashing methods for data referencing and searching files. Open layers are used for overlaying content for virtual data fusion. The prototype uses modular software development to lower costs and open standards for construction pages to improve interoperability. The AW has been demonstrated and tested in a number of exercises.						
15. SUBJECT TERMS						
16. SECURITY CLASSIFICATION OF:				17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 11	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified				

1.0 Introduction

New capabilities are needed to bring together information that is relevant to specific users across a wide range of data sources and media formats. In particular, capability needs have been expressed in Information Technology (IT) requirements of net-centric Programs of Record (PORs). Program Executive Officer for Command, Control, Communications, Computers, and Intelligence (PEO C4I) provides an overview of Science and Technology (S&T) needs [1]. Program Manager, Warfare (PMW) 120 develops, acquires, provides and sustains Information Dominance (ID) capabilities. They have need for S&T to address Intelligence, Reconnaissance, and Surveillance (ISR) in several areas, including storing, accessing and archiving large data sets. They also support Information Operations (IO) in the area of workflow analysis and data fusion. PMW160 desires S&T necessary to handle the rapid growth in data quantity and operations in a cloud environment.

Several military systems under development, such as the Distributed Common Ground Station (DCGS) family of systems, share information across knowledge domains. Many programs are moving to a cloud environment where data/services are delivered over networks. This movement presents both opportunities and challenges. Opportunities are created to acquire a broader range of data and services to support a more complete understanding of the environment. Challenges appear regarding the use of larger quantities of data, specifically use outside of its original intended purpose. Our approach to the problems uses a mixed initiative with both machine automation and human intervention.

The authors' philosophy is that users often do not know in advance what information is needed to perform their work, as discovery is part of the process, and should not be forced to use rigid templates. Users also may not know how to analyze content outside of their domain. Thus, the approach taken was to provide users with situational awareness and the ability to navigate through the information space. Active Wiki (AW) is a system designed for collection, manipulation, collaboration, and rendering of information. AW offers flexible: 1) Information Displays, 2) Tool Interfaces, and 3) Ad Hoc Analyses.

AW permits analysts the ability to pose and experiment with new relationships that may not be well defined using the available data or core ontology. To this end, AW not only contains visualization tools, but also semantic tools for inferencing and consistency checking of data stored in an internal triple-store repository. Automatically collected data and manually entered data are stored in the AW's internal database as flat files, hyperlinked text, or tagged content Resource Description Framework (RDF). Wiki pages are built on-the-fly from the RDF triple-store collection and are dynamically updated. Derivative pages, such as a query output page, can be static and time stamped or set to automatically update when content changes. Similarly, alert pages can be time stamped and saved or set as perishable.

2.0 Background

The wiki paradigm has become a common means to share information both on public domains (Internet) and private domains (Intranet). Research was done to extend this capability to support military users in need of near real time situation assessment. The AW innovations are: 1) "active" monitoring of data sources to automatically build and update wiki pages, 2) lexicon and flexible ontology to build page and entity relationships, 3) open-source software and pathway for a cloud environment, 4) a metadata layer to enable handling of large quantities of diverse data, 5) customizable means for presentation, and 6) a web browser interface for multiple platforms. A summary of the capabilities is provided in Table 1.

UNCLASSIFIED

Table 1. Capabilities explored in AW prototype

Capabilities	Description
Data Folders & Links	Domain lexicon and ontology
Page Content & hyperlinks	Use of concept maps for layout
Maps & Annotation	Metadata for handling big data
Images & Video tags	Metadata for handling big data
Entity recognition	Automatic target recognition
Knowledge Discovery	Filters and graph-based search
Situation Alerts	Event flags in time and space
User specific workspace	Support for user tasks/ missions
Collaboration	Shared workspaces/ annotations
Device Specific Displays	Support for mobile users

The system collects a wide range of data and stores it in wiki pages in a DokuWiki framework. Where possible, the content is semantically tagged using RDF. Open standards, such as Web Services (WS), allow for resource sharing; modular software widgets are compliant with Ozone Widget Framework (OWF), and algorithms facilitate efficient operations in a distributed cloud environment, such as services relying on Apache Hadoop and MapReduce. A summary of software resources is shown in Table 2.

Table 2. The software resources used for AW prototype

Software	Reference
Dokuwiki	http://www.dokuwiki.org/dokuwiki
Fedora Linux 12 x86_64	http://fedoraproject.org/
Apache HTTPD 2.4.2	http://www.apache.org/
PHP 5.4.3	http://www.php.net/
MySQL 5.1.44	http://www.mysql.com/
Ruby-1.9.3	http://www.ruby-lang.org/
Jena 2.6.2 / TSC	http://semanticweb.org/
OpenLayers	http://www.openlayers.org/
Web Map Service	http://www.openlayers.org/
Android OS 2.3 -> 4.0	http://www.android.com/

The AW provides a means for custom handling of data to meet specific user needs as shown in Table 3. AW contains several capabilities in a single package: a semantically tagged wiki with RDF triple-store; tools for interfacing data using SPARQL queries or RESTful web-services; ‘gardening’ tools for examining the semantically tagged content in the wiki; high-level language tool (Ruby) for custom analysis; multidimensional hashing for efficient metadata access; a means to alert users to content or conditions in their location, time, or tasking through mobile devices.

Table 3. The AW built in features and functions

Built-in Feature	Function
Tagging & RDF triple-store	Fusion and inferences for collaboration
Tools for Consuming Data	SPARQL queries or RESTful WS
Inference & Gardening tools	Examining the semantically tagged data
Ruby & High level languages	Ad hoc analysis and visualizations
Multidimensional Hashing	Efficient metadata for fast access
Data Filters & Alert Rules	Filter based on location, time or tasking
Tools for Graphing	Active Graph for SNA of relationships
Tools for Mapping	Active Map for geo-location displays

AW provides the ability to plug in software modules to extend its capabilities, as shown in Table 4. The AW supports recognition of entities and events of interest; interfaces to external databases, messages and maps; and sensor and report collection using built-in or attached sensors. A key challenge is maintaining module integration where output of one module can serve as input to another, thus enabling workflow. This requires that the data be well defined and interchangeable, that services be discoverable, and that performance metrics be used to monitor system operations.

Table 4. The AW external software tested

Plug in Feature	Function
Entity Recognition	Rapier Ship ATR capability
Event Recognition	Activity Discovery using CEP
Databases	GHUB API for resources
Messages	NMCI API for email
Maps & Images	Google Earth maps and imagery
Sensor Collection	ISR App for mobile devices
Report Collection	Software Agents for Websites
Workflow	Software for monitoring workflow

3.0 Overview

Distributed knowledge building is an inherent property of wikis, making it an ideal framework for knowledge sharing and collaborative decision-making. The AW appears as a typical wiki (shown in Figure 1). However, it has advanced features as follows: **Knowledge Framework** for dynamic handling of large data; **Design Concept** for construction and reuse of content; **Interoperability** of application services; and **Visualization** structure where content is stored and can be viewed within a hierarchy of wiki pages. The AW initially provides general data to users based on area, time, and interests. It enables users to explore data and refine collections. Over time, the wiki becomes tailored to better meet the needs of the users.

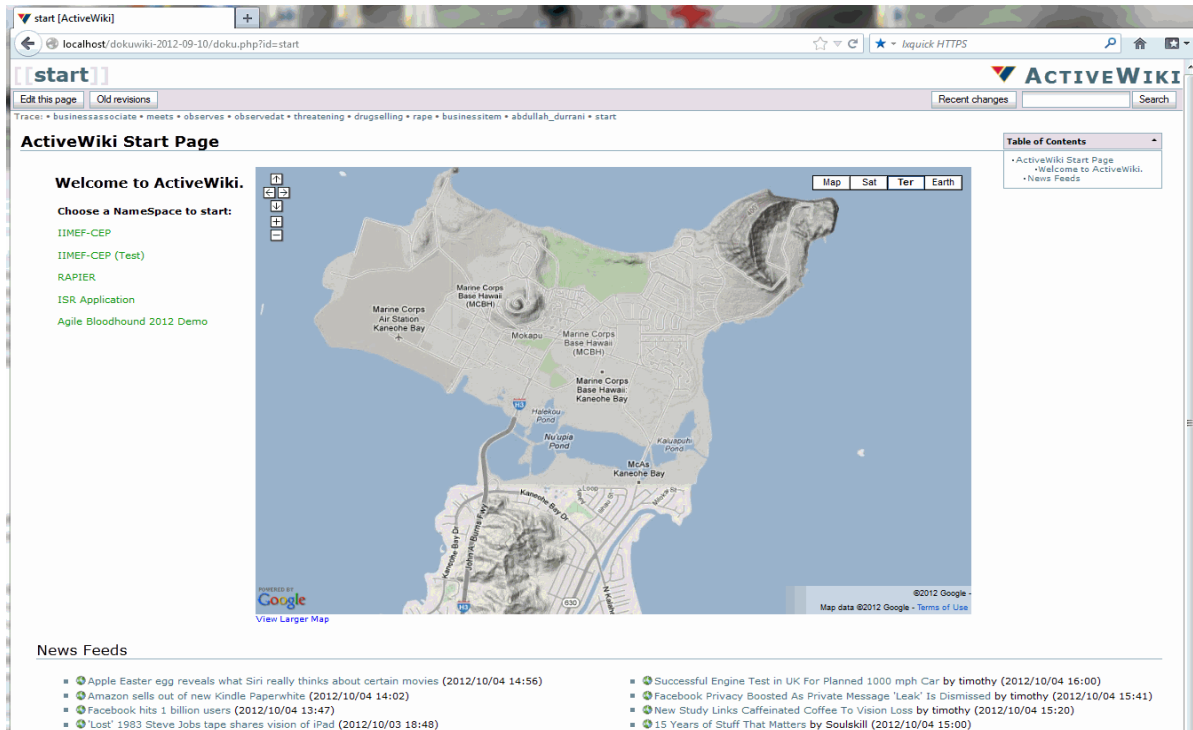


Figure 1 Active Wiki start page appears as typical wiki

3.1. Knowledge Framework

AW utilizes DokuWiki templates and HTML markup to construct visualizations of RDF data extracted from the central RDF triple store. Because the visualizations are coded entirely within Dokuwiki (DW), they may be easily changed to represent needed views as requirements change over time. Furthermore, because data is semantically tagged, data may be further manipulated and visualized directly within the AW, providing further capability to construct customized displays as requirements change. The complete content can be visualized using Active Map. This can be viewed as a form of a concept map for exploring knowledge and gathering and sharing information [2]. Concept maps can be shown as objects or entities (nodes) and relationships (edges) indicated by a connecting line. Products can take the form of a graph, graph with hyperlinks, or website pages [3].

AW also provides the ability for web-enabled applications to provide their own rendering for visualizations requiring more detailed or different visualizations. Examples include Google Earth, Google Maps, blogs, or analyst tool-specific controls or visualizations, in addition to any SOAP-enabled web service. Any tool that can be accessed using a URL in a web browser, or a URI, can be embedded in the AW and the results embedded in AW pages. To take advantage of parallel processing, advanced users would need to setup a workflow (business rules) to partition and assign data and processing. This can be done using Hadoop and MapReduce open standards [4].

3.2. Design Concept

Most knowledge stores build user interfaces to databases. The AW design is unique in that it becomes a knowledge base through composition of its pages. AW exploits semantic tagging of cached content to produce visualizations at several levels of detail. At the ‘lowest level’, data are contained within AW using an ontological superset of a baseline and satellite ontologies. Strictly speaking, AW does not enforce ontological rigor, because it is expected that field analysts would use the AW to enhance the content of the wiki over time, providing the human perspective, collaboration facility, and user commentary. As such, manually entered data is retained separately from automatically generated data and does not risk destruction as the result of automated processes.

Unless otherwise noted, AW operates asynchronously and on-demand with all other services. By design, the contents of AW pages are regenerated on-the-fly from cached contents or data from separate services (web services or central RDF triple store). The only exception to this is web-based embedded applications (See **Interoperability**), which may provide their own synchronization logic whenever that particular Uniform Resource Locator (URL) is being viewed. Because of the asynchronous nature of AW, the order of operations between pages, templates, queries, and transcluded elements may be important in subtle ways. To transclude is to institute a programming step of substituting a template or other input for its rendered text, such as when parsing wiki text [5].

3.3. Interoperability

AW can ingest and render data represented in any of four forms: 1) Embedded display/ user interface described fully by a URL, i.e. hosted by a separate web server (java plugin or javascript); 2) Compose and/or transclude user interface components offered by separate SOAP-style WS ; 3) Internal rendering of RDF data provided by AW from the central store or other stores using AW SPARQL queries and rendering templates; and 4) Interactively share maps and other content using annotation tools to post notes or draw over product displays.

Of the four forms, the first described above provides the greatest control over the display of application service data; it is strongly recommended that application services provide as much of their own rendering as possible or make heavy use of the core ontology. Note that it is possible to crosslink between any and all of the three interface styles as long as the URL/URIs are provided and documented, which would be particularly useful for fusion and entity extraction services. The stream structure of the Knowledge Base Acceleration (KBA) corpus allows the topic entities to evolve. This means that a relevant document from later in the stream might not have been relevant to that entity earlier in the stream [6].

Of the four forms, the second described above provides the second most control of the display of application data. The service produces pre-rendered interface displays (HTML, graphics, etc) that can be rearranged as required. This is customizable through user selection of templates or scripting code. Scripting code is accessible to power-users and developers within the wiki, with no separate development system or platform required, thus simplifying updates. Scripting is “safe”; crashes and loops are caught.

3.4. Visualization

AW can interface to a data store such as Geospatial Hub (GHUB) used by the Marine Corps. GHUB is a distributed content management systems that contains files and links to databases of images, maps, and documents. It contains a number of viewing resources such as CISView, Dropfolder, Falcon View and Google Earth [7]. All data are indexed textually, spatially, and temporally. Full access and control over the system and products is provided using Web Services. AW can ingest data from the GHUB data store as RDF compliant with the ontology described in the referenced document. Data are imported by either

UNCLASSIFIED

exploiting the Data Import extension of DokuWiki automatically, by direct WS, or by using a SPARQL query from embedded code, depending on the need. For simplification, not all data or structure may be present.

The AW is originally designed to meet general tasking or mission needs of a data analyst and user. In the future, this may be done using concept maps to tailor information in a layout for users. Concept maps provide a graphical representation of the data that is a natural fit for data relationships such as a wiki. In addition, these maps can be made consistent with how users learn. Research is needed in the area of shaping a knowledge repository to dynamically adapt it to a user needs. Steps in this direction are being taken in the commercial market with personalized website layout (logons and content preference); tailored advertising (ads specific to users stored on cookies); and online markets (individuals who looked at "A" also looked at "B").

As shown in Figures 2 and 3, AW content can be displayed as graphs or on annotated map/image overlays. The graphs are user interactive; nodes and links can be hidden, highlighted to emphasize, or grayed to reduce notice; and the corresponding AW page will be displayed in a pop-up during mouse-over actions. Annotations can range from simple text to HTML, including other Active Wiki pages as shown, which are fully interactive.

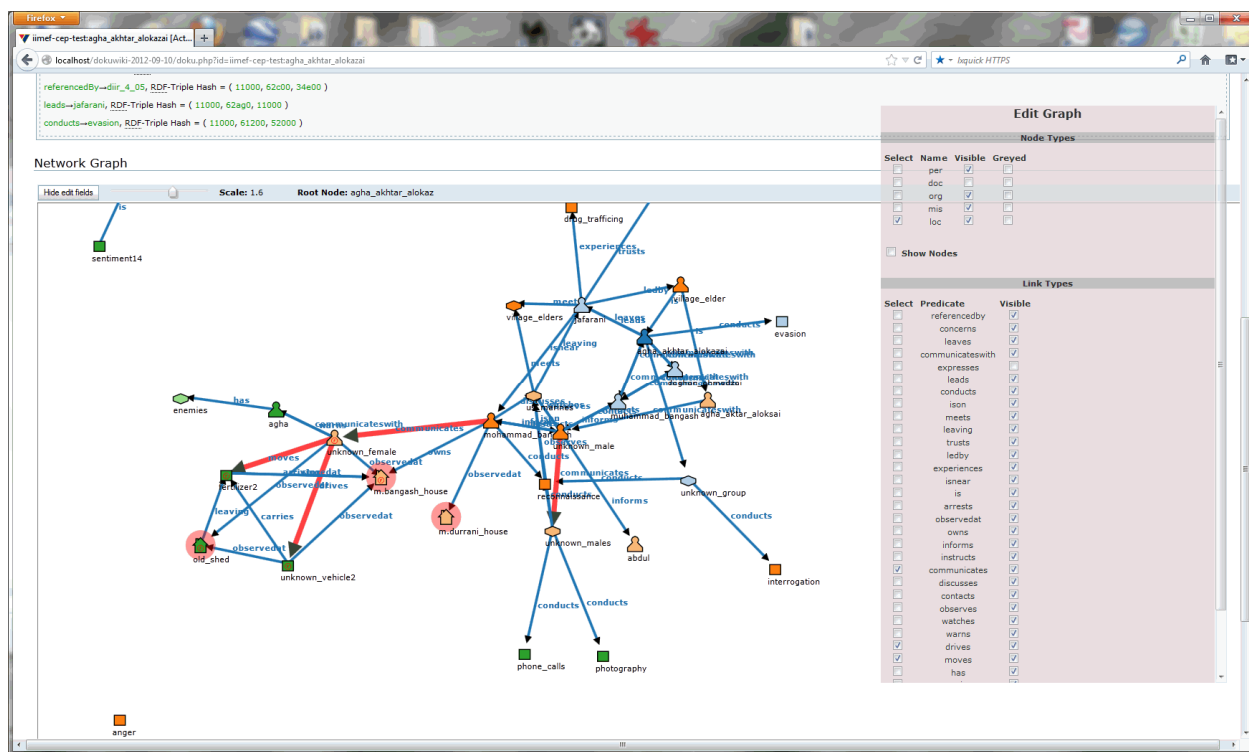


Figure 2 Active Graph that provides visualization and control of pages

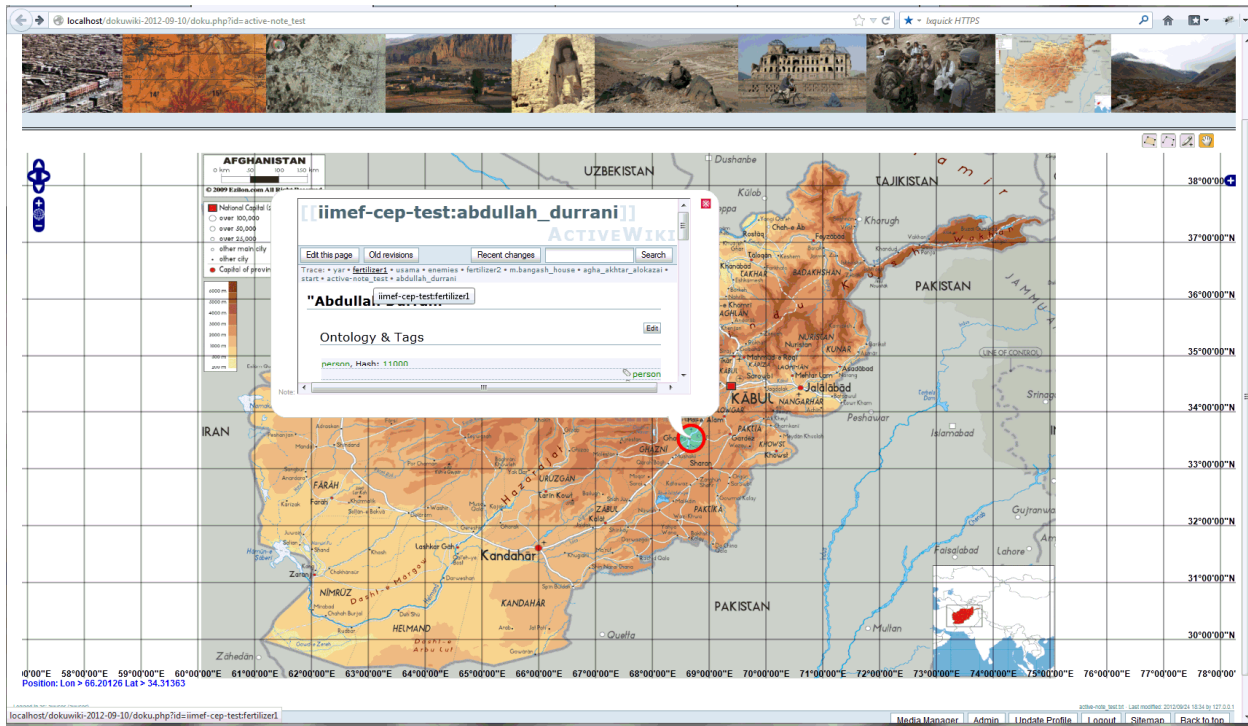


Figure 3 Active Note image display and content annotation

4.0 Data Fusion

The goal of data fusion is to create and maintain a coherent, consistent description of a dynamic situation based on limited observables from one or more sources [8]. Data fusion has been defined as a formal framework in which are expressed the means and tools for the alliance of data originating from different sources [9]. Thus, AW can be considered a framework that supports both the goal and definition of data fusion by actively assembling data into a coherent structure and providing tools for analysis and viewing of content. Unlike a conventional wiki, AW uses a core ontology to assemble and filter information. A data fusion challenge is processing large volumes of data. This is being addressed in the AW by use of metadata hashing and use of resource sharing.

4.1. Metadata Hashing

A challenge of all knowledge repositories is handling large datasets. When dealing with very large data stores the size of the metadata itself can be prohibitive. The commercial market of search engines relies on advanced use of web crawlers to index content on the Internet. While this is useful for static data stores, the dynamic nature of the AW doesn't allow this solution. Our approach is to handle large data streams with hashing methods. SSC Pacific has developed novel methods of hashing to speed access to portions of large files. Conventional means of geo-hashing rely on variable file resolutions that can speed processing but suffer from problems of area boundaries [10]. The SSC Pacific method uses multi-dimensional hashing that avoids these problems.

For example, conventional hashing requires searching for location tags, indexed by lat and lon bounded regions, to retrieve all the tags within the region and then filtering those for a area of interest. The multi-dimensional hashing method makes things that are 'close' in multi-dimensional space 'close' in 1-D hash space. The new method has the following advantages: 1) it is highly scalable; 2) ideally suited for Big Data; 3) supports space-time hashing (easily conducts 'within-bounds' or 'overlapping' searches); 4)

easily builds up 'hot' spots by time and space bounds; 5) supports multi-dimensional hierarchical hashing; and 6) supports RDF-triple hash search by semantic content, not individual nodes/edges [11].

4.2. Resource Sharing

A major challenge is getting the right information to the right person in a timely manner. Google has pioneered distributed data storage, parallel processing, and fault tolerance. This capability has been placed in public domain through open standards [4, 12]. These standards shows promise for distributed data fusion with entity recognition using feature aggregation and track formation by sensor registration and line-of-bearing correlation. Nevertheless, moving data and services in tactical environment is difficult. Networks can vary in bandwidth, user loading and reliability due to the environment. This is an important area of research and a current area of investigation by the authors.

5.0 Experimentation

Experimentation in the field was done to test AW with different operators, devices, and data sources. AW is designed for multiple uses, runs on many platforms and takes most forms of data. It is anticipated that analysts will have a desktop computer system with one or more screens. However, tactical users will often have mobile devices such as laptops, tablets, and smart phones. To accommodate screen-based input (e.g. Android OS) as well as mouse-based input (e.g. Windows OS), the right-click feature of the mouse was not used. In addition, several architectures are possible. For instance, with mobile devices the AW can use servers resident on devices or accessed remotely via networks.

5.1. C4 Applications

AW supports C4 applications normally performed in a command center. The wiki can support both operational and intelligence functions. These services are normally done with desktop configurations. These functions are supported with situation assessment is enabled by active data feeds and displays for specific areas of operation and information requirements. In 2010, AW was used by Marines at the Green Devil portion of Empire Challenge (EC10) exercise at Ft Huachuca, AZ with real-time sensor data. In 2011, AW was used by Marines at IIMEF exercise at Quantico, VA with tactical reports to identify simulated activities. In 2012, AW was used at Agile Bloodhound (AB12) exercise at MCB Kaneohe, HI to both show sensor feeds and analyze data using SSC Pacific products. Real data was used for proof-of-concept testing but was transposed in space and time to fit the AB scenario. Rapier was used for Automatic Target Recognition (ATR) of ships. CEP prototype was used for recognition of activity from TACREP and Draft Intelligence Reports (DIRs). These products were developed separately from AW but are described in the literature [13, 14].

5.2. ISR Applications

SSC Pacific has developed an ISR App that runs on Android OS smart phones or tablets. The ISR App can take pictures that are geo-referenced and placed on a map with augmented feature descriptions. In addition, the user can add notes and affix them to a location as an icon. Shown in Figure 3 is a display at the Tactical Network Topology (TNT) exercise at Camp Roberts Aug 9-11, 2010 [15]. A reconnaissance trip was recorded with a smart phone and displayed on a laptop. The trip was made with a HMMWV that traveled from McMillan airfield to Range 37 as Special Operations Forces operators took pictures en route [16]. The transfer of data can be done by cable or wireless means. The AW has the ability to filter content based on geographic area, tasks and/or time.

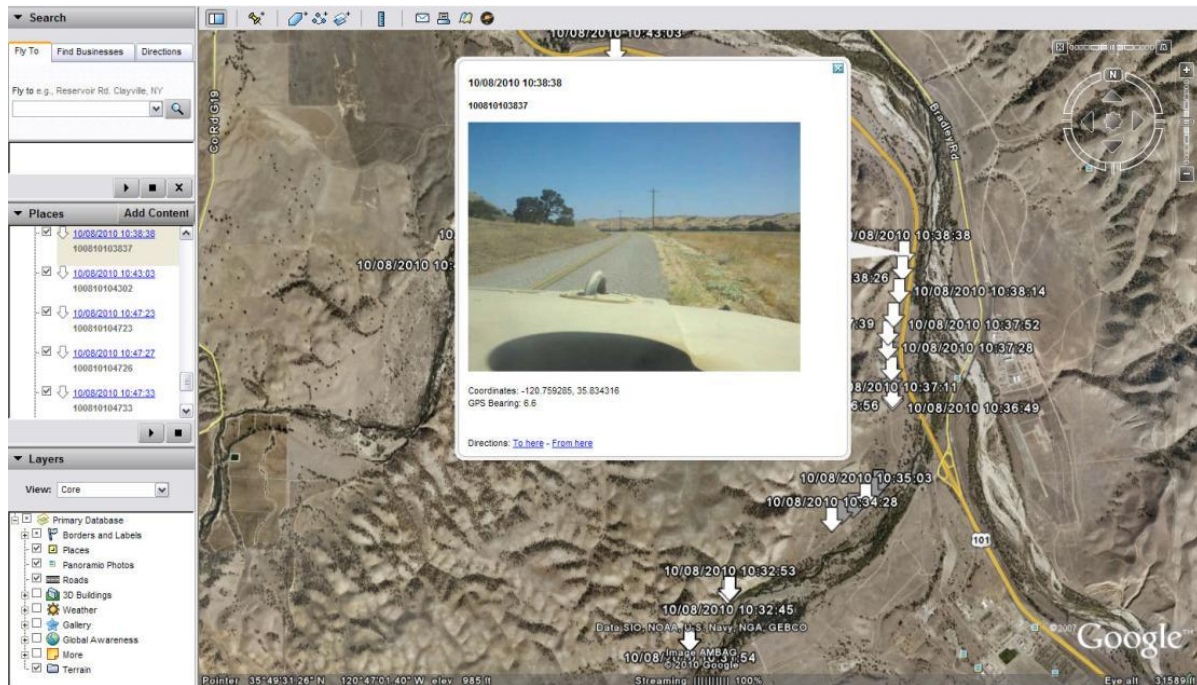


Figure 4 AW displays content based on tasks, area, and time.

6.0 Conclusions

Research on the AW project was done for advanced concept development and demonstration. It has provided insight into the formation of an active knowledge repository. Knowledge generation was achieved not only by algorithmic analysis of external data sources but through active selection of data sources and intelligent composition of content into wiki pages and page relationships. The prototype was used to engage operators and gain insight into needed capabilities. The research led to new data handling and visualization methods. It also served to identify new challenge areas. Many of these challenge areas were written as Small Business Innovative Research (SBIR) topics [17-20].

7.0 Acknowledgement

The authors wish to thank the Office of Naval Research (ONR) for financial support and guidance. They acknowledge and thank the following SSC Pacific scientists and engineers for technical support: Bryan Berg with collaboration tools; Andreas Hirschmann for mobile device ISR App development; Dr. Jim Law for CEP software and assistance in testing; Heidi Buck for Rapier software; Dr. Harry MacKay for workflow software and script development; and Joan Kaina for project support and editorial assistance with the paper. They thank Franklin White (MITRE) for paper presentation.

8.0 References

- [1] Robert Parker and Fernando Escobar, PEO C4I S&T Gaps Overview brief, Mar 12, 2012.
- [2] Joseph D. Novak and Alberto J. Cañas, "The Theory Underlying Concept Maps and How To Construct and Use Them", Institute for Human and Machine Cognition, 2006.
- [3] Plotnick, Eric, "Concept Mapping: A Graphical System for Understanding the Relationship between Concepts", ERIC Digest, 1997. <http://www.ericdigests.org/1998-1/concept.htm>
- [4] Apache Hadoop and MapReduce: <http://hadoop.apache.org/>
- [5] Wiktionary definition of transclude, <http://en.wiktionary.org/wiki/transclude>
- [6] TREC Knowledge Base Acceleration, 2012 <http://trec-kba.org/kba-ccr-2012.shtml>
- [7] GHUB brief by NRL-Stennis, 2012.
- [8] White, Franklin E., Data Fusion Subpanel Report SIGINT Session in Technical Proceedings Volume 1 of 1990 Joint Service Data Fusion Symposium, 15-18 May 1990.
- [9] Wald L., 1999. Some terms of reference in data fusion. IEEE Transactions on Geosciences and Remote Sensing, 37, 3, 1190- 1193.
- [10] Geohashing, Wikipedia, <https://en.wikipedia.org/wiki/Geohash>
- [11] Wroblewski, Ronald, Multidimensional hashing, patent pending.
- [12] Accumulo, <http://accumulo.apache.org/>
- [13] Jim B. Law and Scott C. McGirr, "Multi-INT Complex Event Processing using Approximate, Incremental Graph Pattern Search", 17th ICCRTS, Washington DC, June 19-22, 2012
- [14] Bryan Bagnall, Elan Sharghi, and Heidi Buck, Algorithms for the Detection and Mapping of Wildfires in SPOT 4 and 5 Imagery, Submitted to SPIE Conference Proceedings, 2012.
- [15] TNT Exercise, sponsored by SOCOM and NPS.
- [16] Capabilities Based Experimentation (CBE) Quick Look Report (QLR) "Knowledge Discovery using Smart Phones", Aug 11, 2010.
- [17] SBIR OSD09-SP6 Semantic Wiki for Page Alerting (2009)
- [18] SBIR OSD09-SP4 Designing Large Data Handling Architectures (2009)
- [19] SBIR N102-176 Disambiguation of Entity Association Statements (2010)
- [20] SBIR OSD10-L06 Knowledge Discovery Using Mobile Devices (2010)